Hamilton Standard Division
United Aircraft Corporation
Windsor Locks, Connecticut

First Quarterly Progress Report - NASw-871

I   Introduction

28939

The over-all goal of this project is the development and application of statistical decision techniques appropriate to decision problems involving the selection and implementation of complex, costly biological experiments, especially those characterized by relatively high degrees of uncertainty regarding the statistics of the processes to be observed and unusual degrees of complexity (and concomitant unreliability) of the observation equipment to be employed. It was decided at the suggestion of NASA Headquarters personnel that the efforts of the project be applied to the extraterrestrial life detection program, since the selection and implementation of experiments associated with this program are good examples of the class of decision problems with which this project is concerned. The activities of the first quarter were directed primarily at surveying some of the decision problems associated with this program, orienting project personnel, identifying concrete subproblems to be employed as vehicles for the development of operational analysis and statistical decision techniques, and initial efforts directed toward exploring and solving these problems. A list of trips made and persons contacted, and a brief description of the first model to be simulated are appended.

II   Project Activities and Accomplishments

A.   Structuring of General Problem of Extraterrestrial Life Detection

The method chosen for the development of statistical decision techniques is an empirical one--that is, our procedure is to take concrete problems and attempt to solve them, developing and testing solution techniques in the process. Given this decision regarding the method of approach, the first and most critical step is to structure the general problem of extraterrestrial life detection in such a fashion as to permit the selection of appropriate subproblems which are amenable to operational analysis and the application of statistical decision techniques.

The task of structuring a problem of this nature is essentially the task of identifying and eliminating from consideration those problems which, for one reason or another, are not worthwhile, either because of practical or theoretical limitations inherent in the mathematical approach; or because the recommendations resulting from the application of the mathematical evaluation techniques cannot, in fact, be implemented. The most acute theoretical limitations have to do with the difficulty of comparing alternatives whose utility is not expressible in common terms such as money, time or probability of failure. It is a characteristic of complex problems such as the extraterrestrial life detection problem that many of the decisions which must be made involve factors which, although tangible, cannot be compared. Other types of problems may be in principle amenable to mathematical analysis but are so complex as to involve an exhorbitant and wholly impractical amount of labor. For example, it is easy to define stochastic models which would take years to program and tens of years to evaluate on the fastest modern computers. In addition, some evaluation procedures may in principle be satisfactory, but may, in fact, be inapplicable because they assume the existence of facts, theories, or policies which do not exist. And, finally, implicit in the notion of evaluation is the assumption that the courses of action being evaluated are real ones which could actually be carried out.

The task of structuring a general problem may be seen to consist of becoming sufficiently familiar with the practical problem so as to break it down into sub-problems, and to distinguish subproblems which can usefully be attacked from those which cannot. In the process of doing so one also develops an estimate of the difficulty of the appropriate subproblems and tentatively identifies solution techniques. Another desirable outcome of the structuring activity is the determination of the probable importance of each of the subproblems.

Most of the efforts of the first quarter has as their objective the structuring of the general problem of extraterrestrial life detection in such a form as to permit the identification of concrete subproblems to which our effort should be directed. This consisted of a partial survey of the status of the program, conducted primarily by means of visits with NASA personnel at Headquarters and Ames Research Center, and with NASA contractors (Jet Propulsion Labs, Hazelton Laboratories, Stanford Medical Center, etc.). On each of these visits we interviewed those contacted about their efforts in the Exobiology program and solicited their opinion as to what represented the outstanding problems associated with the extraterrestrial life detection program, and with each we discussed possible approaches to these problems. These visits were supplemented with literature surveys. Although the most significant result of these activities consisted in the definition of concrete problems described in the next section, certain tentative conclusions of a more general nature were made and are summarized below:

1. Many of the participants in the program are acutely concerned with the problem of defining the concept of "life" or at least with developing an explicity list of "life characteristics" which can be employed in evaluating alternative experiments. No such definition currently exists, and an acceptable definition may never be developed. (The need for one may be exaggerated.) Therefore, we should not consider any evaluation procedures which assume the existence of a single, widely accepted definition of this nature.

2. The most readily accessible and useful subproblem areas appear to be those concerning questions of implementation of experiments, as these seem to be the only ones for which relatively hard facts are available and in terms of which realistic alternative courses of action can be defined. Simulation of such experiments with the objective of evaluating alternative implementations or alternative payload mixes appears to be the most fruitful direction for this project.

3. The most useful mode of operating appears now to consist of participating in joint evaluation efforts with the individual experimenters, relying on them to guide us to practical problems and to provide much of the input data and interpretation of results.

4. Most of the experimenters contacted have devoted relatively little attention to the problems of interface between the unit for which they are responsible and the rest of the system, and consequently system interface specifications are lacking or unrealistic. Since the manner in which these details are implemented may be of crucial importance, it would seem desirable to explore the consequences of various "interface" effects at as early a date as possible. Since the final design will consist of a compromise between scientific and engineering considerations, it is important that the scientists remain cognizant of the effects of engineering compromises which might be made. Simulation appears to be a way to explore some of these consequences prior to the actual integration phase.

5. There is a need to compare experiments with respect to their "value" in providing evidence for or against the existence of life on Mars. Many schemes now being considered seem to assume either agreement as to the defining character- istics of life, or greater knowledge than is now available about the environment of Mars or about the significance re the occurrence of Martian life of chemical or physical properties of Mars. All such schemes would appear, at least for our purposes, to suffer from the lack of agreement by scientists in the relevant fields. Alternative bases for comparison may, in fact, exist and should be looked for.

The task of structuring the problem is not complete in that we expect that additional subproblem areas will continue to present themselves as our knowledge of the program increases and as the program itself evolves.

B. Problem Definition

Two subproblem areas were identified. The problem areas and our proposed solution techniques are described below together with a summary of the reasons that were selected for study.

1. Data Transmission Requirements - The first problem to which project efforts have been directed is that of determining the amount and quality of the data which should be transmitted from each experiment in order to minimize the probability of ambiguous or misleading conclusions based on those data. The major considerations which lead to the selection of this as a subproblem are summarized below:

1. Although the amount of information—in bits per second—which can be tele- metered back to earth depends upon the size of the booster and the payload employed; it does not appear now that any of the contemplated missions will have bandwidth to spare. Mariner-type missions will have very little com- munication capability since they will depend on direct transmission, and increasing bandwidth will require allocations of limited power and payload weight to communications rather than to the operation of the experiments themselves. Current plans for larger vehicles which will land larger pay- loads utilize orbiting busses for relaying information back to earth, but the demands on the communication channel are greater in that more experi- ments are planned.

2. Relatively few experimenters have been able as yet to specify their data transmission requirements. In some cases this is due to the status of the experiment itself; in others, it is also in part due to the fact that the mission profile has not yet been determined and the interface details have not been specified.

3. Alternative allocations of channel capacity to different experiments are possible, and consequently evaluations of various of these alternatives would seem to provide a useful guidance. Similarly, the designs of each of the experiments could be modified if it appeared that alternative methods of collecting and transmitting the data were advantageous.

4. Preliminary hand simulation of a model of an experiment conducted by us indicated that some experiments might be highly vulnerable to bandwidth limitations. (This simulation is described in more detail below.)

5. Most of the experimenters whom we contacted have considered data trans- mission requirements only to the extent of specifying the minimum number of bits required to transmit a significant message under the assumption

5. that all components of the system function as designed. Under such conditions a minimal number of "readings" of the instrument will suffice to transmit useful information. As yet, relatively little consideration appears to have been given to the consequences of malfunctions or to other possible sources of misleading data.

Hand simulation of a model experiment indicated that allocation of channel capacity to transmission of information regarding the function of components may greatly reduce the probability of drawing erroneous conclusions, and this suggests that such alternative should be evaluated.

The technique to be employed in determining the amount and quality of data to be transmitted is essentially the development and evaluation of stochastic models of the various experiments. (The simulation technique is described in more detail below.) An appropriate model will be designed and programed, and variations in the frequency of observations, the number of bits per observation message, the amount of status information transmitted, sensitivity and error proneness of the discriminators will be compared. Only as much of the experiment as seems appropriate to this objective will be modeled in detail. The first experiment which will be examined in this fashion is Gulliver. A preliminary analysis of Gulliver components that may be modeled has been made.

2. Techniques for Comparing Different Experiments - The second subproblem area which has been identified is that of comparing different experiments with respect to their utility, so that decisions regarding the allocation of limited amounts of resources can be made appropriately. It was pointed out in Section A above that the need to compare experiments with respect to their "value" in providing evidence for or against the existence of life of Mars is widely recognized. Although many procedures for comparing the utility of the various experiments are, in principle, definable, most of them appear to assume the existence of an accepted explicit definition of "life" and/or more knowledge about Mars or about the relationship betwen discoverable physical properties and the occurrence of life than is now available. Such assumptions are, therefore, not valid at this time, and evaluation procedures based on them do not at present seem worthwhile. Nevertheless, resource allocations must be made, and for this reason it has been decided that some project effort should be expended in the search for techniques by which certain aspects of the alternative experiments may be compared. A brief consideration of the types of resource allocations which must be made suggests that many of them will concern the selection of a mix of experiments to be included in a single payload and of the allocation of weight, space, power, etc. to the various experiments. Although considerations of the relative scientific value of each of the possible experiments will be highly relevant to such decisions, numerous other factors which can be more readily evaluated will also be relevant. Some project effort has been devoted to identifying factors which can be more readily evaluated and to exploring ways in which alternative experiments can be compared with respect to these factors. Certain of these factors--availability, readiness for incorporation into a payload, and over-all cost--serve to limit the set of experiments which may be considered for a given launch date, but do not provide a basis for selecting among alternative allocations of resources, or for determining the extent to which a given experiment enhances the value of a payload. Additional factors which may provide a basis for more subtle comparisons have been tentatively identified. Most of them concern the interpretability of the information provided by the experiments and the mutual interactions of the experiments. Four of the most promising comparison factors are briefly described below:

1. Availability of standard for interpretation of the data. The experiments now under consideration appear to differ widely in so far as the interpretability of the possible outcomes. For example, an almost infinite number of infrared spectra can be generated, but most of these cannot be interpreted, in that it is rarely possible to infer from a given spectrum the identity of the compounds or mixtures of compounds from which those spectra were obtained. This suggests that it might be worthwhile to explore whether experiments may be usefully compared in terms of the ratio of the number of interpretable outcomes to the number of possible outcomes.

2. Reliability and Error Rates. In **conventional** statistical hypothesis testing, recognition is given to two types of error, viz, the rejection of the null hypothesis when in fact it is true, or acceptance of the null hypothesis when indeed it is false. These are the errors of the first and second kind. The probabilities associated with committing these errors are denoted respectively by α and β .

It can be shown that for a given sample size N:

$$\alpha = f(\beta)$$

(1)

where $f(\beta)$ denotes a more or less complicated function.
In conventional experimental programs it is customary to fix α and β at some preselected level and then search for <u>experimental designs</u> such that **N** is as small as possible.

As is shown in books on statistics, the theory of hypothesis testing in general concerned with the construction of tests using data in which the "true" value of the variable of interest in subject to a random "error." Thus, instead of being able to observe μ which for the sake of illustration we might imagine as the actual number of a certain species of microorganisms in a gram of soil, we can only observe:

$$x = \mu + e$$

(2)

where e is a random fluctuation caused by **unknown** or uncontrollable sources of variation.

Now, if we make many repeated observations and collect a <u>sample</u> of x and if the average of such a smaple approaches μ as the size of the sample gets large, then:

$$E(x) = E(\mu + e) = \mu + E(e) = \mu$$

(3)

where operator E is the "expection" operator, and we say that x is an unbiased estimate of μ . In other words, the average value of the fluctuations e is zero.

This type of model works moderately well under the restrictions normally imposed in conventional experimental situations in which extensive equipment calibrations and checks can be made to insure adequate operation of the measuring devices.

In considering the operation of remote, automatic equipment, a third source of error becomes significant. We shall call this the error of the third kind, namely the data were obtained by partially or totally **malfunctioning** equipment. Errors arising from malfunctioning equipment differ from the fluctuations conventionally considered in that:

but

$$E(x) = \mu + b$$

where b is a (positive or negative) bias.

As mentioned above, in conventional experimentation means can usually be found to eliminate equipment failure as a source of erroneous data. Consequently, the probability of an error of the third kind can be set to zero. In automatic experimentation it is clear that this probability cannot be equated to zero even approximately.

We must, therefore, distinguish in some quantitative manner between "usable" observations which are contaminated by random error, but which tend to yield estimates which are accurate although they may be unprecise, and "unusable" observations which are corrupted by bias due to partial malfunction and which will result in inaccurate estimates (which may, of course, be quite precise). To eliminate bias due to partial or complete malfunction, one must increase the reliability of the system as much as possible and then provide for a means of identifying the occurrence of malfunctions so that as many as possible of the "unusable" observations can be identified as such. The liability to detected and undetected bias obviously affects the over-all desirability of an experiment, since data known to be biased are usually useless, and biased data not recognized as such result in inaccurate estimates. It may be assumed that all experiments are made as reliable as practical in that they are constructed of components which have a low probability of failure. Two experiments which are equally reliable in that they are both as reliable as possible, may in fact be effectively different in that they differ with respect to the likelihood of detected and undetected bias and with respect to the effects of the undetected bias on the probabilities of inaccurate conclusions. If liability to detected and undetected bias can be estimated, then a comparison of experiments in terms of the consequent proneness to erroneous estimates (due to undetected bias combined with actual random errors) may be feasible.

3. Demands on System. Some experiments may be characterized as "greedy" in that they require relatively large amounts of scarce commodities such as power, space, communication channel capacity, etc., or require special control features or protected in-flight environments. The extent to which any experiment is a liability is a function of the extent to which it degrades the performance of companion experiments, or reduces the over-all reliability of the system. Specification of system demands and liabilities, if practical, would permit the comparison of experiments with respect to their effects on specifiable subsets of potential companion experiments.

4. Complementary and Redundancy. It has been suggested that one rule to employ in the selection of a subset of experiments is that the experiments should be complementary in that they observe characteristics of Mars or of Martian life that are likely to occur independently of one another. Implicit in this strategy is the notion that if two different experiments are designed to detect the same property (or properties which may be presumed to co-occur), then if that property is not present neither experiment will return any useful information, whereas if that property is present the experiments return the same information and one of them is redundant. The reasoning behind the complementary rule is that we do not know which of a very large variety of possible states of Mars will be encountered by the probe. If the probe consists of experiments which are complementary, then the likelihood that

4. the state encountered will be characterized by at least on property which the probe can measure is greater than would be the case if the experiments were redundant. This rule and the strategy implicit in it would appear to be a very sensible one. It is, however, one which we believe may be difficult to apply. To answer the question of whether or not two given experiments which measure superficially different properties are or are not complementary, one must know something about the probably co-occurrence of the properties in question on Mars. Thus, the answers which one gets may depend upon which scientist one asks. Another difficulty with this rule is that implicit in its employment is the assumption that all of the experiments result in the acquisition of accurate unambiguous estimates of the parameters in question, for two experiments which are, in fact, redundant will give the same answers only if they both function perfectly. Considered from this point of view, it is apparent that no two experiments are, in fact, redundant. If the experiments are "different" then they measure the same property but in different ways, and hence are subject to different kinds of errors and liable to different kinds of bias. Even if the experiments are replications of the same experiment, the amount of bias and sampling error represented in a single run of each may differ, even to the extent of resulting in apparently inconsistent observations. Similarly, it will be observed that two experiments which appear to be complementary in that they yield observations of properties which do in fact occur independently, may, nevertheless, be dependent on one another in the sense that they may both be vulnerable to one or more partial or complete failures due to system malfunction.

In this sense, a set or mix of experiments may be said to exhibit varying degrees of independence with respect to a given system topology. Clearly, when several experiments are to be "packaged" in one vehicle a **maximum** degree of system independence is desirable in order to reduce the vulnerability of each experiment to the fate of any other portion of the mission.

C. Simulation.

Computer simulation of stochastic models of life detection experiments is the
technique which will be employed to explore the data transmission requirements of
Gulliver and other selected experiments. Much of the first quarter efforts were
devoted to developing models and simulation techniques and are itemized below:

1. Preliminary Models and Hand Simulations -- An extremely abstract and highly
simplified model of an experiment involving the culturing of a sample of Martian
soil was designed and a few hand simulations were conducted. The variables which
were varied in the few hand simulations conducted were frequency of observations,
duration of experiment and "telemetry" error. The results of these hand simulations
are not included here, but may be summarized as supporting the not surprising con-
clusion that the ability to draw accurate unambiguous conclusions from the over-
simplified experiment modeled varies with all three of these variables. It was
therefore decided that more sophisticated models should be designed and programmed
for computer simulation.

2. Soup Gedanken -- A more sophisticated and somewhat more realistic model
of a life detection experiment was designed and programming of this model for
evaluation on a PDP-1 computer was initiated. Soup Gedanken is an experiment in-
volving the collection and culturing of a sample of Martian surface. It is an
imaginary experiment in the sense that it is a fairly abstract and general purpose
model that bears little direct resemblance to any of the experiments currently under
development. The purpose of designing and evaluating Soup Gedaken was to: (1) pro-
vide a vehicle for the development of necessary computer routines (especially input
and output routines) that can be used in more realistic special purpose simulations
of experiments, (2) provide a debugged "sample" simulation which can be used to
demonstrate to biologists some of the kinds of trade-offs which can be explored by
means of simulation techniques, and (3) to examine in a more accurate way the kinds
of effects which frequency of observation, number of bits per message, and total life
of the experiment may have.

Soup Gedanken is designed as a game which may be "played" by the "experimenter."
The experimenter "designs" the experiment by specifying the values of certain para-
meters in terms of which the controllable variables are represented in the computer
program. The experimenter specifies a family of growth curves by drawing a curve
on the face of a CRT which is read by the computer. He also specifies the duration
for which the experiment is to be run (which may vary from 16 to 128 experimental time
units) and the frequency with which observations are to be made (every time unit,
every other time unit, etc.). Virtually all portions of the program are expressed
as functions of variables to which different values may be assigned at different runs,
and in this sense virtually all of the variables may be controlled by the "experimenter."

Having made these assignments, the experiment is conducted by means of the execution
of sub-routines simulating various modeled components. These include the following:

1. A density function whereby the number of organisms in the innoculum varies
with the landing site.

2. The growth rate function which determines the speed of growth of organisms,
the duration of the lag preceding population increases, the generation time which is
a function of the curve provided by the experimenter, and the size of the innoculum
and three random variables.

3. A colony generation function which utilizes the output of the growth rate function, but modifies it by adding random perturbations corresponding to the effects of uncontrolled variables such as fluctuations in temperature.
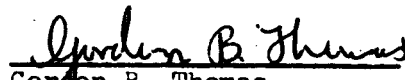
4. A "discriminator" which corresponds to the device which measures the quantity of an accumulated metabolic product (represented as the sum of a constant times the population at each experimental time interval). The discriminator is assumed to have a threshold corresponding the minimal amount of product which must be present to obtain a non-zero output, and a saturation level corresponding to the quantity which results in the maximum reading. The number of discriminative intervals and a discrimination error are also modeled.

5. A "telemetry model" which determined the number of bits which should be transmitted for each observation and reverses some of the actual bits in accordance with a preset probability of bit reversal. The final corrupted messages are stored (together with the number and magnitude of each telemetry error.).

Having run the experiment and generated the messages, the telemetered data points are then displayed to the "experimenter" who must decide whether any Martian organisms were cultured. He can replay the experiment as many times as he wishes, changing the total duration of the experiment and/or the frequency of observations. After having examined the messages he may inquire into what actually occurred, and obtain displays of the population counts, the discriminator output, and a relevant print-out.

Soup Gedanken is partially programmed and debugged and several modifications to the original model have been designed. The program may also be run in an automatic mode rather than a game mode so that large numbers of simulated experiments can be conducted in sequence and statistics regarding the performance of the modeled experimental apparatus can be generated.

Dian R. Hitchcock

Gordon B. Thomas

SITE VISITS

PERSON	DATE	LOCATION
Dr. R. Young Mr. H. Klein	3/24/64 "	Ames, Moffet Field, Calif. "
Dr. E. Levinthal	3/25/64	Stanford, California
Mr. G. Hobby	3/26/64	Jet Propulsion Labs, Pasadena, Calif.
Dr. C. Sagan	4/8/64	Smithsonian Astrophysics Lab Cambridge, Mass.
Dr. O. Reynolds	4/21/64	Washington, D. C.
Dr. V. Oyama	4/30/64	Ames, Moffet Field, Calif.
Mr. G. Hobby Dr. G. Soffen	5/1/64 "	Jet Propulsion Labs, Pasadena, Calif. "
Dr. W. Vishniac Dr. E. Weston	5/11/64 "	University of Rochester, New York "